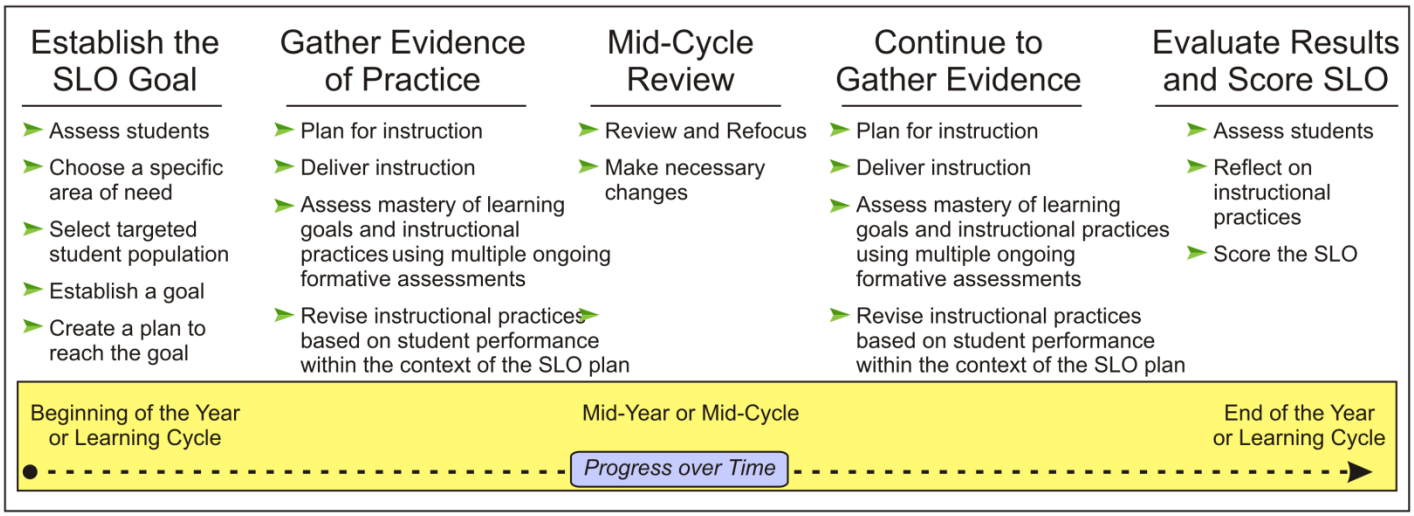
# Using a Balanced Assessment Framework to Support the SLO Process

The purpose of this document is to demonstrate the way that a balanced assessment framework can support the Student Learning Objectives or the School Learning Objectives (SLO) process for educators, and to review examples of commonly used assessments that may aid an educator and an evaluator in setting and measuring SLO processes.

## *The SLO Process and a Balanced Assessment Framework*

The SLO process consists of four main elements spread out over a school year or learning cycle:

- Establishing an appropriate SLO goal and plan to reach the goal
- Gathering evidence of instructional practices leading to improved student outcomes
- Conducting a mid-year or mid-cycle review of progress
- Evaluating the final results and scoring the educator's SLO progress and outcomes




Educators use a variety of assessment tools to gather data about student performance when establishing an SLO goal. There are three different "families" of assessment tools in a balanced assessment framework:

- **Benchmark (or Interim) Assessments**: Periodic diagnostic or progress assessments that benchmark and monitor progress.
- **Formative Assessments**: Daily, ongoing evaluations that quickly and immediately inform instructional practices that support student learning.
- **Summative Assessments**: Large scale standardized assessments that evaluate cumulative student learning.

Every assessment tool has a specific intended use for measuring student learning. Determining the best assessment tool to use depends on aspects of the need, such as:

- The **specificity** of data needed
- The **kind** of data needed
- The **timing** of the assessment or amount of time between assessments
- The **frequency** of reassessment for the same information

A critical aspect of the SLO process is to evaluate how the decisions, practices and strategy choices that an educator makes in planning for and delivering instruction ultimately affect student progress toward a goal. Multiple measurements and assessments used as part of a balanced assessment framework are beneficial in helping educators triangulate data, validate practices, and support informed choices that lead to increased student outcomes.

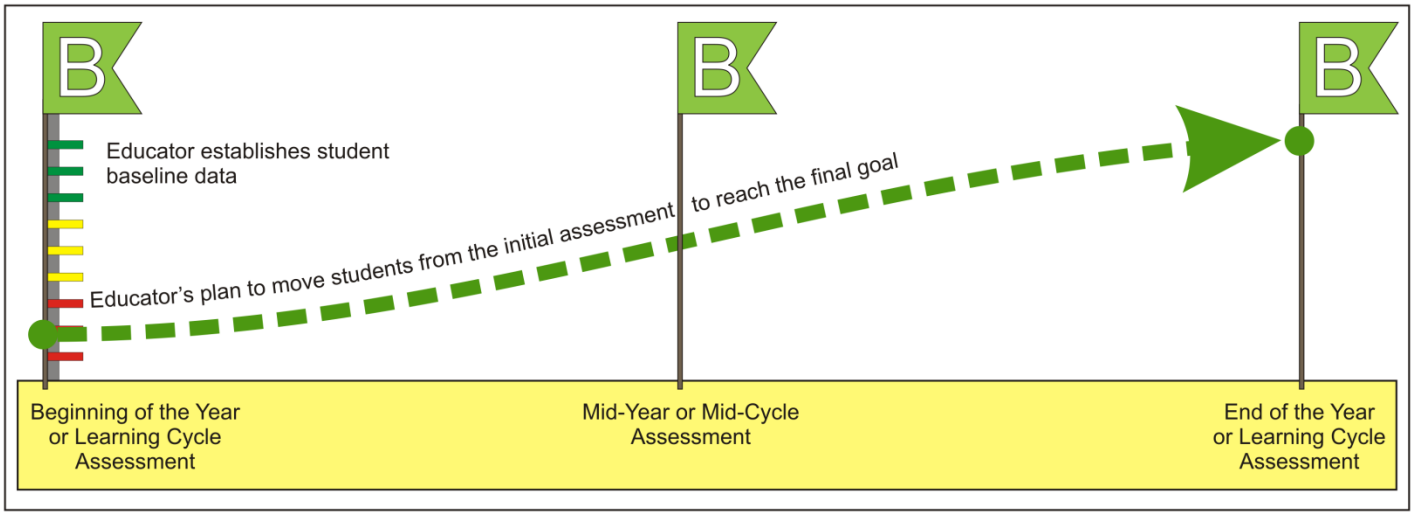
It is also important to remember that how an assessment tool is used can change what kind of assessment tool it is. For instance, an assessment commonly used 2-3 times a year as a benchmark (interim) assessment (ex. Measures of Academic Progress-MAP) could become a summative assessment if it was only used once a year, at the end of a school year, to measure the growth from the previous school year to that point.

The scope of what the assessment tool measures must match the amount of instruction or skills being assessed. For instance, formative assessment tools are intended to be used frequently and to assess fairly small amounts of progress, or to assess student mastery of smaller skills that support progress toward the larger goal. Summative assessments measure complex sets of skills or learning over a longer period of time.

## *Starting Strong: Using Benchmark Assessments to Support the SLO*

A typical SLO process starts with three main pieces: first, identifying a specific group of students based on a baseline assessment of student skills; second, identifying a desired skill, growth or achievement goal for the targeted student population; and third, identifying a plan to guide the students to reach the goal. Frequently, schools or teachers will use **benchmark (or interim) assessments** at key points in a learning year or cycle to know how a student or school is able to perform prior to or after receiving direct instruction.

**Benchmark (or interim) assessments** are assessments that are administered periodically, generating multiple data points across time, usually 2-3 times a year or during a learning cycle (for instance, beginning – middle – end, or fall – winter – spring). They are classroom or school-centered in scope, and are used to "benchmark" or determine a student or school's current status as it relates to a goal. Occasionally, different or multiple benchmark assessments are used to assess the same skill.




Educators use benchmark assessment tools at the beginning of the learning cycle to gather baseline data on a student's current performance, and then use the same tool or a comparable tool at the end of the cycle to evaluate the growth made. Often the educator will use the tool at or around the mid-cycle point as a way of knowing if the student is on track to grow. Using the same or comparable assessments gives the educator the ability to know where a student starts and then select a goal point on the same scale.

For example, a 1st grade educator may administer the Fountas and Pinnell assessments to determine a student's reading level at the beginning of the year. The student may demonstrate that they read at the "C" level, but the teacher knows that to be on target to perform at norms, that student needs to reach level "J" by the end of the year. The teacher may set short term goals to get the student to Level "G" by the semester, and level "J" by the end. The educator would re-assess the student's reading level, and make adjustments to his or her plan based on the student's progress.

**Examples of commonly-used benchmark (interim) assessments include:**

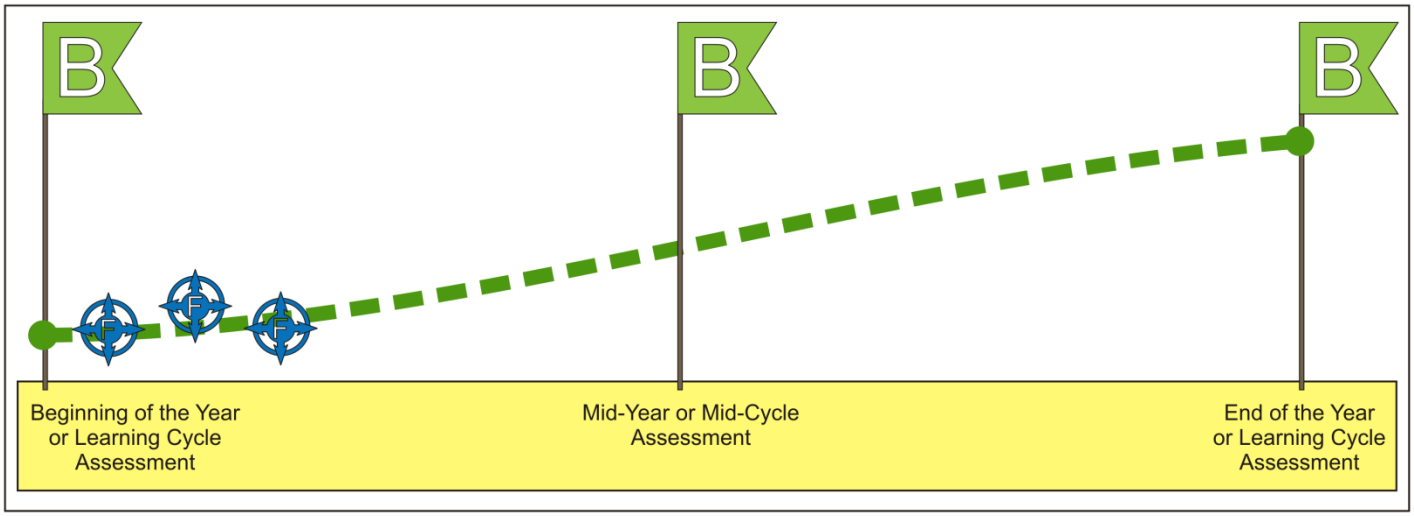
| Assessment Name | Grades | Administration | Use of the Data |
|---|---|---|---|
| Ages & Stages | 4K | Approximately 30 minutes, administered week prior to the start of the school year | Data is used at the start of the 4K experience to determine any developmental delays for district response and early intervention. |
| PALS-Phonological Awareness Literacy Screening | 4K – 1st | Approximately 30 minutes, administered Fall and Spring | Data is used to identify students in need of additional reading instruction and informs instruction for all young readers. |
| DIBELS-Dynamic Indicators of Basic Early Literacy Skills | K – 6th | Approximately 3-10 minutes per probe, Benchmark probes are administered Fall, Winter, and Spring; Progress monitoring probes can be administered as frequently as weekly or monthly | Data from probes is used to screen students for intervention and provide teachers ongoing progress monitoring for students receiving intervention.<br><br>Note: DIBELS probes can be used as a formative assessment. |
| NWEA MAP-Measures of Academic Progress (Common Core Version) | 2nd – 8th, infrequently used with HS grades | Mathematics, Language Usage and Reading Goals Survey Tests, approximately three hours; administered in Fall, Winter, and Spring. | Data is used as an instructional tool for informing instruction, creating flexible groups, identifying students for services or interventions, and placing students. |
| AIMSweb | 4K – 12th | Probes range in time to administer (1-10 minutes each); Benchmark probes are administered Fall, Winter, and Spring; Progress monitoring probes can be administered as frequently as weekly or monthly | Data from probes is used to screen students for intervention and provide teachers ongoing progress monitoring for students receiving intervention.<br><br>Note: AIMSweb probes can be used as a formative assessment. |
| STAR-Renaissance Learning Reading, Math, and Early Literacy | 4K – 12th | Approximately 15-20 minutes per module; Benchmark probes are administered Fall, Winter, and Spring; Progress monitoring probes can be administered as frequently as weekly or monthly | Data from probes is used to screen students for intervention and provide teachers ongoing progress monitoring for students receiving intervention.<br><br>Note: STAR modules can be used as a formative assessment. |
| SRI-Scholastic Reading Inventory | 3rd – 8th | Approximately one hour; administered at the beginning and end of instruction. | Data is used to inform reading instruction and provide services or interventions. Data can also be used to match texts to literacy level (Lexile) |
| Odyssey Explorer-Compass Learning | K – 8th | Approximately one hour each for reading, language arts, and math; administered up to three times each year. | Data is used as an instructional tool for informing instruction, creating flexible groups, identifying students for services or interventions, and placing students. |
| Smarter Balanced Assessment | 3rd – 8th, 11th | 2 assessments per grade: one computer adaptive and one computer-based performance task. | Summative accountability standards and reporting, ensuring curriculum alignment to the standards.<br><br>Note: Some components of Smarter Balanced may be used as interim (benchmark) or formative assessments. |
| ACT Aspire Early High School | 9th – 10th | | Data is used for plotting growth to the ACT assessment, summative accountability standards and reporting, ensuring curriculum alignment to the standards.<br><br>Note: ACT Aspire assessments are more predominately used as a summative assessment. |

## *Providing Instruction: Gathering Data and Evidence of Practice*

Benchmark (interim) assessments are integrally helpful to an educator to know if they are on the path to achieving the goal, but they are only guideposts. To use an analogy, benchmark assessments are helpful in the same way that a person traveling from Milwaukee to St. Paul by car knows that they are generally on the right path because they have passed next to both Madison and Eau Claire on the way.

The limitation of benchmark assessments is that they don't directly reveal how choices made during instruction have impacted instruction (positively or negatively) on a more immediate scale. To gain that understanding, educators use **formative practices**.

**Formative Practices** are the instructional strategies that educators use to assess student understanding on an ongoing basis. Sample strategies include spot questioning, on the fly quizzes, or discussion between student and teacher in which the teacher assesses a student's or group of students' understanding of the learning objective. Formative practices are used at the beginning, during and at the end of a lesson. They are frequent, no-stakes "check-ins" administered quickly. The results of formative practices tell the educator and student how well the student is understanding the lesson/concept and gives the educator feedback to refine their instruction.




After beginning instruction, educators use formative practices to fine-tune their instructional choices and plans. Because they can be used frequently, educators may use multiple formative assessments to monitor progress between benchmarks. The formative assessments provide critical feedback that helps the educator refine their approach and keep the student moving consistently toward the target. Benchmark and summative assessments are often thought of as assessments *of* learning, while formative assessments are assessments *for* learning.

**Examples of commonly-used formative assessments include:**

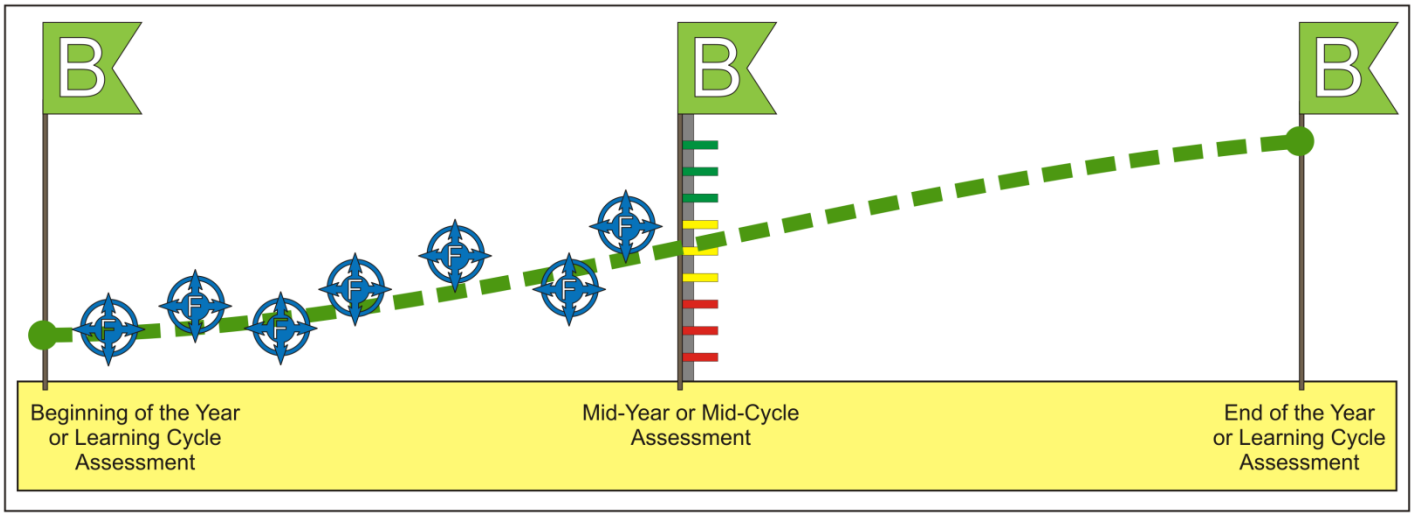
| Assessment Name | Grades | Administration | Use of the Data |
|---|---|---|---|
| Teacher-created materials | 4K – 12th | Variable length of time; teacher created and administered usually during less than one full class period. Could be paper, performance, or technology based. | Data is used to determine progress toward a goal with specific skill sets that support a larger goal. |
| Fountas and Pinnell Leveled Reading Assessment | K – 5th | Approximately 10 minutes, administered three times a year (K-2) or as needed (3-5) | Data is to determine individual reading levels in order to inform instruction. |
| Odyssey-Compass Learning Custom Assessments | K – 8th | Variable length of time; teacher created and administered electronically | Data is used to determine progress toward a goal with specific skill sets that support a larger goal. |

**Examples of commonly-used formative assessments (continued):**

| Assessment Name | Grades | Administration | Use of the Data |
|---|---|---|---|
| AIMSweb | 4K – 12th | Probes range in time to administer (1-10 minutes each); Benchmark probes are administered Fall, Winter, and Spring; Progress monitoring probes can be administered as frequently as weekly or monthly | Data from probes is used to screen students for intervention and provide teachers ongoing progress monitoring for students receiving intervention.<br><br>Note: AIMSweb probes can be used as a benchmark assessment. |
| STAR-Renaissance Learning Reading, Math, and Early Literacy | 4K – 12th | Approximately 15-20 minutes per module; Benchmark probes are administered Fall, Winter, and Spring; Progress monitoring probes can be administered as frequently as weekly or monthly | Data from probes is used to screen students for intervention and provide teachers ongoing progress monitoring for students receiving intervention.<br><br>Note: STAR modules can be used as a benchmark assessment. |
| DIBELS-Dynamic Indicators of Basic Early Literacy Skills | K – 6th | Approximately 3-10 minutes per probe, Benchmark probes are administered Fall, Winter, and Spring; Progress monitoring probes can be administered as frequently as weekly or monthly | Data from probes is used to screen students for intervention and provide teachers ongoing progress monitoring for students receiving intervention.<br><br>Note: DIBELS probes can be used as a benchmark assessment. |
| UW Early Placement Exams | 10th – 12th | Approximately 2-3 hours, administered throughout the year independent of the local district | Data is used to determine progress toward achieving a 3, 4, or 5 on an AP exam or informing college placement. |

## *The Mid-Cycle Review: Pause to Reflect, Refocus, and Reevaluate*

Around the halfway point in the learning cycle (semester, year, etc.), an educator may again administer a benchmark assessment. Benchmark assessments can be used again to get an idea of how much progress a student has made using the same scales or assessments that were used to generate the baseline data.



At the mid-cycle point, the educator has two important sets of data upon which to draw: first, two sets of benchmark assessment results; and second, reflection on and documentation of the formative practices that they have used to guide student learning

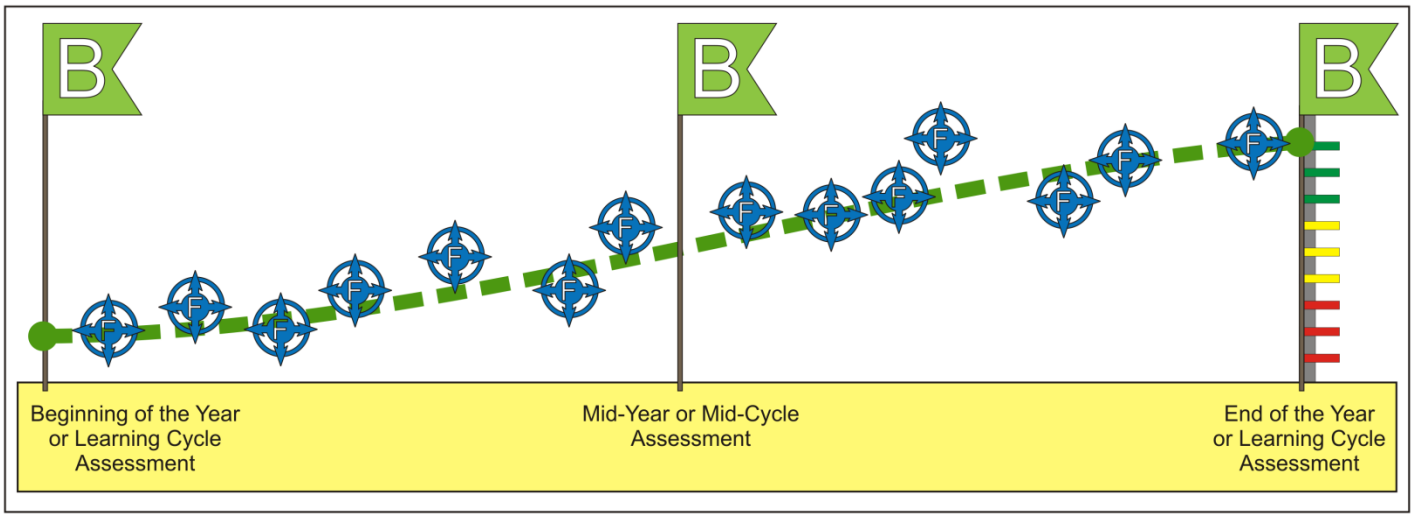
Using these two data sets to study student progress can help an educator evaluate the choices they have made and revise any aspects of the SLO plan that may need revision. For instance, the educator could discover:

- The SLO goal was set too low initially, as the students have almost reached the end-of-cycle goal by the mid-cycle point.
- The SLO goal was set too high initially, as the goal seems to be out of reach even though the students are making adequate progress.
- Students should be added to or removed from the targeted group.
- The educational strategies that the teacher is using aren't as effective as they should be, because students aren't making adequate progress on pace to the goal, but the goal is clearly appropriate.

The mid-cycle point is the time when an educator and his or her evaluator meet to discuss progress and make revisions so that the SLO plan can generate the best possible student outcomes.

## *Continuing Forward: Gathering Additional Data and Evidence of Practice*

After the mid-cycle review, educators continue to gather evidence of student progress through the use of formative practices. Over the course of a school year or learning cycle, a typical path with multiple practices used along with ongoing formative and benchmark assessments may look like this:



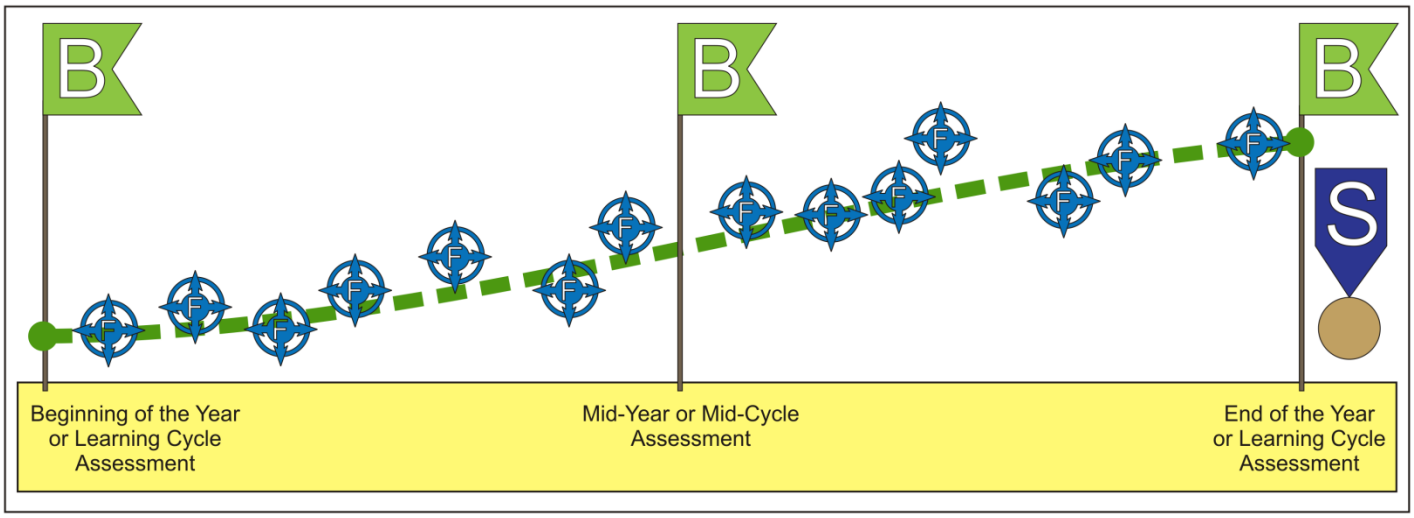

**When scoring the SLO, the evaluator will want to see the documentation of an educator's formative practices that demonstrate how an educator guided their students to reach their goal.** The conversations between an educator and an evaluator will center on showing evidence of what an educator has done, rather than evaluating what their students have achieved strictly through assessment data. In this manner, assessments support SLO's by helping an educator make the best choices throughout the process to generate the best possible student outcomes.

When ongoing formative assessment data shows that the educator's practices and decision-making is yielding student results that are on the line or above the line, the educator knows that they are making the right kind of progress and their choices are validated. When formative assessment data shows that the student results are falling below the line, it tells the educator that they need to change their approach or strategies.

## *Reaching the End of the SLO Process*

At the end of a year or learning cycle, the student or school is often measured for proficiency or mastery using a **summative assessment**.


**Summative assessments** are large-scale (relative to instruction) assessments that are given very infrequently, and measure overall student growth. Often, summative assessments are used to measure multiple skills at the same time, and are not very useful in crafting individual instruction strategies. Item analysis of summative assessments is valuable in evaluating the alignment of curriculum to standards on a larger scale.




The final assessment may tell the educator if their students or school hit their goal, but will not reveal what strategies or practices the educator used to move the students or school along the path. That information will need to be gathered throughout the cycle of instruction, and be used along with a framework of assessments to demonstrate to what degree the educator's practices can be attributed to the students' or school's growth.

Ultimately, the totality of the evidence an educator presents during the collaborative scoring process informs the conversation. Summative assessment achievement data may provide important triangulation and validation of the results from benchmark and formative assessments. By the time the end of the year or cycle has arrived, it should be fairly easy for an educator to present evidence of their impact by showing what they did to support the instructional goal. It should also be very easy to use this evidence to clearly reflect on their practices and draw conclusions that inform their practices in the future.

**Examples of commonly-used summative assessments include:**

| Assessment Name | Grades | Administration | Use of the Data |
|---|---|---|---|
| Teacher-created tests and examinations or school/district common summative assessments | 4K – 12th | Variable; depends on the level, the subject, the goals and the instruction. | Data is used to assess mastery of a unit's worth of study. |
| WKCE-CRT | 3rd – 8th, 10th | Gr. 3: 4 hours, 15 min.<br>Gr. 4: 6 hours, 35 min.<br>Gr. 5: 4 hours, 20 min.<br>Gr. 6: 4 hours, 35 min.<br>Gr. 7: 4 hours, 35 min.<br>Gr. 8: 6 hours, 55 min.<br>Gr. 10: 6 hours | Data is used for federal accountability and creation of school report cards.<br><br>Item analysis of student performance is used to inform instruction and ensure curriculum alignment to state standards. |

**Examples of commonly-used summative assessments (continued):**

| Assessment Name | Grades | Administration | Use of the Data |
|---|---|---|---|
| Smarter Balanced Assessment | 3rd – 8th, 11th | 2 assessments per grade: one computer adaptive and one computer-based performance task. | Summative accountability standards and reporting, ensuring curriculum alignment to the standards.<br><br>Note: Some components of Smarter Balanced may be used as interim (benchmark) or formative assessments. |
| ACT | 10th – 12th | Approximately 4 hours at a testing center, usually outside of the school day. Starting 2014-15, all 11 grade students will take the ACT during the school day. | Data is used to support College or IHE entrance |
| SAT | 10th – 12th | Approximately 4 hours at a testing center, usually outside of the school day. | Data is used to support College or IHE entrance |
| ACT Aspire Early High School | 9th – 10th | | Data is used for summative accountability standards and reporting, ensuring curriculum alignment to the standards.<br><br>Note: ACT Aspire can be used as a benchmark assessment. |
| AP Exams | 11th – 12th | Approximately 3 hours at the student's school, during the school day, on prescribed dates. | Data is used to demonstrate mastery of advanced placement coursework and can impact College or IHE entrance and placement |
| ACT Work Keys | 11th | 3 assessments, approximately one hour each | Data is used to assess career and workforce readiness, and supports the Career Readiness certificate program |

## *Integrated systems of assessment and instruction*

There are a relatively new set of tools available to districts and teachers that both assess student learning and provide adaptive instruction within the same electronic framework. Adoption of these systems has become popular as a means to differentiate instruction and provide intervention strategies within a Response to Intervention (RtI) framework. Many of these systems are intended to act as short-term, intensive systems that take a laser-like focus on skill development and help students to close achievement gaps. Other systems are used to support larger instructional initiatives, such as demonstrating proficiency with technology or supporting a 1:1 computing initiative.

**Examples of commonly-used integrated assessment and differentiated instruction systems include:**

| Assessment Name | Grades | Administration | Use of the Data |
|---|---|---|---|
| ALEKS-Assessment and Learning in Knowledge Spaces | K – 12th | Approximately one hour per subject level, predominately math and math-based science; Benchmark assessments | Data is used as a tool for informing instruction, creating flexible groups, identifying students for services or interventions, and placing students. |
| PLATO and Study Island with Edmentum Assessments | K – 12th | Approximately one hour per subject level, core, electives, career/technical ed, world languages, AP; Benchmark assessments | Data is used as a tool for informing instruction, creating flexible groups, identifying students for services or interventions, and placing students. |
| Read 180 and the Scholastic Reading Inventory (SRI) | 4th – 10th | Approximately one hour for the SRI, less time for weekly probes, literacy and comprehension based; Benchmark assessments | Data is used as an instructional tool for informing instruction, creating flexible groups, identifying students for services or interventions, and placing students. |

**Examples of commonly-used integrated assessment and differentiated instruction systems include:**

| | | | |
|---|---|---|---|
| Achieve 3000 | 2nd – 12th | Approximately one hour per subject level, predominantly literacy and science; Benchmark assessments | Data is used as an instructional tool for informing instruction, creating flexible groups, identifying students for services or interventions, and placing students. |
| TAGLIT, Atomic Learning, and Learning.com | | 21st century skills and technology literacy | Data is used to support technology literacy skill mastery and support technology implementation plans, including 1:1 initiatives |